DOCUMENT RESUME

ED 391 834 TM 024 435

AUTHOR Jo, See-Heyon
TITLE Multilevel Analysis of Structural Equation Models via the EM Algorithm.
PUB DATE [95]
NOTE 19p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Algorithms; *Bayesian Statistics; Equations (Mathematics); *Error of Measurement; Estimation (Mathematics); *Maximum Likelihood Statistics; Sampling; *Statistical Bias; *Structural Equation Models
IDENTIFIERS *EM Algorithm; Hierarchical Models; Missing Data; *Multilevel Analysis

ABSTRACT

The question of how to analyze unbalanced hierarchical data generated from structural equation models has been a common problem for researchers and analysts. Among difficulties plaguing statistical modeling are estimation bias due to measurement error and the estimation of the effects of the individual's hierarchical social milieu. This paper presents empirical Bayes estimation by means of the EM algorithm in the context of unbalanced sampling designs. The EM algorithm is particularly useful when analytic expressions exist for the conditional distribution of the missing data given complete data and for the complete-data maximum likelihood estimates of the model parameters. One example is included for illustrating the specification of the model that incorporates variations at two levels of hierarchy. (Contains 2 tables, 1 figure, and 32 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 391 834

Multilevel Analysis of Structural Equation Models via the EM algorithm

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

See-Heyon Jo

College of Education
Michigan State University

PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
SEE-HEYON JO
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Abstract

The question of how to analyze unbalanced hierarchical data generated from structural equation models has been a common problem for researchers and analysts. Among difficulties plaguing statistical modeling are estimation bias due to measurement error and the estimation of the effects of hierarchical, social milieu where individuals are situated. This paper presents empirical Bayes estimation by means of the EM algorithm in the context of unbalanced sampling designs. The EM algorithm is particularly useful when analytic expressions exist for the conditional distribution of the missing data given complete data and for the complete-data maximum likelihood estimates (MLE) of the model parameters. One example is included for illustrating the specification of the model which incorporate variations at two levels of hierarchy.

Key words: Structural equation model, EM algorithm, hierarchical data.

TM024435

BEST COPY AVAILABLE

1. Introduction

As a consequence of various theoretical developments and of improvements in computing, maximum likelihood (ML) estimation has become a viable procedure for estimating parameters in multilevel structural equation models under the balanced sampling design. The initial interest in ML estimation of a multilevel covariance structure model was noted and developed by Schmidt (1969). Schmidt and Wisenbaker (1986) extended this work to the structural equation models (Joreskog, 1973) for balanced hierarchical data.

McDonald and Goldstein (1989) derived the likelihood equations and derivatives for a bilevel structural equation model which allows for variables measured strictly at a higher level, though no computational approach was made available. They also indicated that the procedure for computation of ML estimates is currently less well developed for the unbalanced sampling design.

Recently, based on the balanced-data theory provided by Schmidt (1969) and McDonald and Goldstein (1989), Muthen (1990) showed that the maximum likelihood fitting function could be rewritten such that the between and within structural models could be estimated by means of a multi-population analysis in LISCOMP (Muthen,1987) or other comparable structural equations software. In the case of balanced data this could be accomplished by treating the within-group deviations as sampled from one population and the between-group deviations as sampled from a second population. For the case of

unbalanced data, each cluster of groups which have the same number of observations is treated as one population.

Lee and Poon (1992) also used the strategy of classifying level-2 units into D subsets of level-2 units having equal sample size. They proposed an estimator for such data which, though not maximum likelihood (ML), has the same asymptotic distribution as the ML estimator as the number of level-2 units per subset increases without bound. Computationally this estimator is available using standard software program, such as LISREL (Joreskog and Sorbom, 1993) or EQS (Bentler,1989).

More recently, Raudenbush (1993) proposed an alternate approach for the unbalanced case. He conceptualized the problem in the framework of groups which could all have the same number of sampled cases but are missing data for some individuals. In particular, in the M-step (maximization) the method uses the standard program such as EQS (Bentler, 1989).

This paper presents empirical Bayes estimation procedures for computing ML estimates of the parameters in the multilevel structural equation models under the unbalanced sampling design. The primary merit is that the procedures provide us with ML estimates of the structural coefficients for each level-2 unit (e.g., group). In particular, the EM algorithm is particularly useful when analytic expressions exist for the conditional expectation of the missing data and for the maximum likelihood estimates of the model parameters given the observed data and missing data. The paper presents a multilevel structural equation modeling process with an example, which serves to illustrate the general principles of empirical Bayes estimation with the EM algorithm.

2. The Multilevel Structural Equation Models

A simple measurement equation for each individual is

$$y_{ij} = \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \qquad (1)$$

$$\varepsilon_{ij} \sim N(0, \Sigma), \quad \Sigma \text{ is } Diagonal(\sigma_r^2).$$

where j=1,2,..,J for classrooms, i=1,2,...,n_j for students nested in classroom j, r=1,..., R for latent constructs. The vector of y_{ij} represents the set of observed endogenous measures, and the vector of η_{ij} stands for the latent endogenous variables on the within level, while the vector of η_{bj} stands for the latent endogenous variables on the between level, Λ_w and Λ_b are coefficient matrices relating latent variables and y. The subscript "w" denotes the within-level, while "b" denotes the between-level.

Assuming structural linear relationships among constructs, the theoretical relationships on the within-level can be expressed through the structural equation

$$\eta_{ij} = A_j \eta_{ij} + B_j z_{ij} + u_{ij} \qquad (2)$$

$$u_{ij} \sim N(0, \Delta)$$

Equation (2) stipulates that on the individual-level the latent variables are a linear function of themselves and the exogenous predictor variables measured without error. For example, in equation (2) one of the components of z_{ij} is gender, while the other component takes on a unit value so that the model has an intercept.

Now we reduce equation (2) into

$$\eta_{ij} = (I - A_j)^{-1} B_j z_{ij} + (I - A_j)^{-1} u_{ij} \qquad (3)$$

so that the resulting reduced form for the within-level structural equations (2) is

$$\eta_{ij} = Z_{ij}\pi_j + \theta_{ij} \tag{4}$$

where

$$\theta_{ij} = (I - A_j)^{-1} u_{ij}\ ,\ \theta_{ij} \sim N(0, T_{\eta_j}) \tag{5}$$

$$\pi_j = vec^{*}(I - A_j)^{-1} B_j \tag{6}$$

where vec^{*} stacks the transpose of each row of a matrix into a vector and the matrix A_j is fixed rather than random for each group. The parameters of the structural coefficients for each group, α_j's, are computed from the variance covariance matrix, T_{η_j}. After finding structural coefficients $\hat{\alpha}_j$'s, from equation (6) we find regression coefficients, $\hat{\beta}_j$'s. We now capture π_j as a function of level-2 unit's predictors, W_j.

$$\pi_j = W_j\pi_0 + v_j \tag{7}$$

$$v_j \sim N(0, T_\pi)$$

On the between-cluster level, the structural relationships can be expressed as

$$\eta_{bj} = A_b\eta_{bj} + B_b w_j + u_{bj} \tag{8}$$

$$u_{bj} \sim N(0, \Delta_b)$$

Equation (8) is the expression for the structural relationships among latent variables and the exogenous predictor variable on the group level. All of exogenous predictor variables are observed directly without error, e.g., school location (rural, urban), school sector (public, nonpublic), religion, gender, ethnicity, family size (numbers of a household), individual's age, current membership in a political party or sports club.

Then we have the reduced form for between-level structural equations (8)

$$\eta_{bj} = W_j \pi_{bj} + v_{bj} \tag{9}$$

$$v_{bj} \sim N(0, T_{\eta_b}),$$

By substituting the structural equations (9), (7) and (4) into (1), we have the following combined equation (10). This representation permits us to develop a special version of the EM algorithm for multilevel structural equation model:

$$y_{ij} = [\Lambda_w Z_{ij} W_j | \Lambda_b W_j] \begin{bmatrix} \pi_0 \\ \pi_{b0} \end{bmatrix} + [\Lambda_w Z_{ij} | \Lambda_b] \begin{bmatrix} v_j \\ v_{bj} \end{bmatrix} + \Lambda_w \theta_{ij} + \varepsilon_{ij} \tag{10}$$

In a more compact form we have:

$$y_{ij} = \Lambda_0 \tilde{Z}_{ij} \pi + \Lambda_0 Z_{ij}^{*} \varpi_{ij} + \varepsilon_{ij} \tag{11}$$

where,

$$\Lambda_0 = [\Lambda_w | \Lambda_b],$$

$$\tilde{Z}_{ij} = \begin{bmatrix} Z_{ij} & 0 \\ 0 & W_{ij} \end{bmatrix}, \quad Z_{ij} = \begin{bmatrix} z_{1ij} & z_{2ij} & 0 & 0 \\ 0 & 0 & z_{1ij} & z_{2ij} \end{bmatrix},$$

$$W_j = \begin{bmatrix} w_{1j} & 0 \\ 0 & w_{1j} \end{bmatrix}, \quad Z_{ij}^{*} = \begin{bmatrix} I & Z_{ij} & 0 \\ 0 & 0 & I \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\pi = \begin{bmatrix} \pi_0 \\ \pi_{b0} \end{bmatrix}, \quad \pi \sim N(0, \Gamma), \quad \varpi_{ij} = \begin{bmatrix} \theta_{ij} \\ v_j \\ v_{bj} \end{bmatrix},$$

3. The EM Algorithm for Maximum Likelihood Estimates

Dempster, Laird and Rubin (1977) presented the EM algorithm as a general iterative method for computing maximum likelihood estimates in the presence of incomplete or missing data. Wu (1983) presented its properties of convergence,

viewing it as a special optimization algorithm. The EM algorithm is particularly useful when analytic expressions exist for the conditional distribution of the missing data and for the maximum likelihood estimates of the model parameters given the complete data. By casting the measurement model and the reduced form of the structural equation for latent variables into a special form of the mixed model (Raudenbush, 1988), we can conceptualize our problem as having complete data and incomplete data.

Given

$$
\begin{aligned}
y_{ij} &= \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \\
\eta_{ij} &= Z_{ij} \pi_j + \theta_{ij} \\
\pi_j &= W_j \pi_0 + v_j \\
\eta_{bj} &= W_j \pi_{b0} + v_{bj}
\end{aligned}
\tag{11}
$$

we have $\Theta = \{\Lambda_w, \Lambda_b, \Sigma, T_{\eta_b}, T_{\eta_j}\}$ is the set of unknown parameters, where

j=1,...,J.

$y_{obs} = \{y_{ij}, z_{ij}, w_j\}$ is the set of observed data.

$y_{mis} = \{\theta_{ij}, v_j, v_{bj}, \varepsilon_{ij}, \pi_0, \pi_{b0}\}$ is the set of missing data.

$c = \{y_{obs}, y_{mis}\}$ is the set of complete data.

Note that in our multilevel structural equation model the factor loadings are unknown parameters rather than observed predictors.

The joint probability density function of missing data and observed data are :

$$f(c|\Theta) = (2\pi)^{-Nr/2}|\Sigma|^{-N/2}\exp\Big[(-0.5)\sum\sum(y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi - \Lambda_0 Z_{ij}^{*}\varpi_{ij})^T\Sigma^{-1}(y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi - \Lambda_0 Z_{ij}^{*}\varpi_{ij})\Big]$$

$$\times(2\pi)^{-JP/2}|T_{\eta_1}|^{-n_1/2}\exp\Big[(-0.5)\sum\theta_{i1}^T T_{\eta_1}^{-1}\theta_{i1}\Big] \times,\ldots,\times(2\pi)^{-JP/2}|T_{\eta_J}|^{-n_J/2}\exp\Big[(-0.5)\sum(\theta_{iJ}^T T_{\eta_J}^{-1}\theta_{iJ})\Big]$$

$$\times(2\pi)^{-Jq/2}|T_\pi|^{-J/2}\exp\Big[(-0.5)\sum(v_j T_\pi^{-1} v_j)\Big] \times(2\pi)^{-JS/2}|T_{\eta_b}|^{-J/2}\exp\Big[(-0.5)\sum(v_{bj}^T T_{\eta_b}^{-1} v_{bj})\Big]\times h(\pi)$$

(12)

where r=the number of indicators, p=the dimension of θ_{ij}, q= the dimension of π, and s=the dimension of v_{bj}. The prior density $h(\pi)$ is considered a very small constant and it can be ignored while the empirical Bayes estimators are calculated (Dempster, Rubin & Tsutakawa, 1981).

From Dempster, Laird and Rubin (1977), each iteration of the EM algorithm solves $\frac{\partial}{\partial\Theta}[Q(\Theta,\Theta^{(i)})|_{\Theta=\Theta^{(i)}}] = 0$ for Θ,

where $Q(\Theta,\Theta^{(i)}) = \int L(\Theta|c) f(y_{mis}|y_{obs},\Theta=\Theta^{(i)})dy_{mis}$, and $L(\Theta|c)$ is the complete-data log-likelihood. In the E-step, the conditional expectation of the complete data log likelihood is taken with respect to the posterior distribution of the missing data. The necessary posterior location vector and dispersion matrix of missing data in model (11) is directly derived from the Appendix (Raudenbush, 1988). These posterior distributions given ML parameter estimates provide point estimates and intervals needed for inference about missing data. To find the ML estimates, we have to maximize the function $Q(\Theta^{(i)},\Theta^{(i-1)})$ by taking its first derivatives with respect to $T_\pi, T_{\eta_b}, T_{\eta_j}, \Sigma, \Lambda_0$.

$$(1)\ \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial T_{\eta_j}} = -\frac{n_j}{2}[2T_{\eta_j}^{-1} - D(T_{\eta_j}^{-1})]$$

$$+\frac{1}{2}\sum_{i=1}^{n_j}[2T_{\eta_j}^{-1}E(\theta_{ij}\theta_{ij}^T|Y=y,\Theta^{(i-1)})T_{\eta_j}^{-1} - D\{T_{\eta_j}^{-1}E(\theta_{ij}\theta_{ij}^T|Y=y,\Theta^{(i-1)})T_{\eta_j}^{-1}\}]$$

where D denotes a diagonal matrix (Graybill, 1983). Thus D(T) is a diagonal matrix with i-th diagonal element equal to the i-th diagonal element of a matrix T. Setting the derivative equal to null matrix and solving gives the complete-data ML estimate.

Then the complete-data ML estimate is :

$$\hat{T}_{\eta_j} = \frac{1}{n_j}\left[\sum(\theta_{ij}^{*(i)}\theta_{ij}^{*(i)T} + D_{\theta_{ij}}^{*(i)})\right] \tag{13}$$

$$(2)\ \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial T_{\eta_b}} = -\frac{J}{2}[2T_{\eta_b}^{-1} - D(T_{\eta_b}^{-1})]$$

$$+\frac{1}{2}\sum_{j=1}^{J}[2T_{\eta_b}^{-1}E(v_{bj}v_{bj}^T|Y=y,\Theta^{(i-1)})T_{\eta_b}^{-1} - D\{T_{\eta_b}^{-1}E(v_{bj}v_{bj}^T|Y=y,\Theta^{(i-1)})T_{\eta_b}^{-1}\}]$$

Setting the derivative equal to null matrix and solving gives the complete-data ML estimate. Then the complete-data ML estimate is :

$$\hat{T}_{\eta_b} = \frac{1}{J}\left[\sum(v_{bj}^{*(i)}v_{bj}^{*(i)T} + D_{v_{bj}}^{*(i)})\right] \tag{14}$$

$$(3)\ \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial T_{\pi}} = -\frac{J}{2}[2T_{\pi}^{-1} - D(T_{\pi}^{-1})]$$

$$+\frac{1}{2}\sum_{j=1}^{J}[2T_{\pi}^{-1}E(v_jv_j^T|Y=y,\Theta^{(i-1)})T_{\pi}^{-1} - D\{T_{\pi}^{-1}E(v_jv_j^T|Y=y,\Theta^{(i-1)})T_{\pi}^{-1}\}]$$

Setting the derivative equal to null matrix and solving gives the complete-data ML estimate. Then the complete-data ML estimate is :

$$\hat{T}_{\pi} = \frac{1}{J}\left[\sum (v_j^{*(i)} v_j^{*(i)T} + D_{v_j}^{*(i)})\right] \tag{15}$$

$$(4)\ \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \Sigma} = -\frac{N}{2}[2\Sigma^{-1} - D(\Sigma^{-1})]$$
$$+\frac{1}{2}\sum\sum[2\Sigma^{-1}E(\varepsilon_{ki}\varepsilon_{ki}^{T}|Y=y,\Theta^{(i-1)})\Sigma^{-1} - D\{\Sigma^{-1}E(\varepsilon_{ki}\varepsilon_{ki}^{T}|Y=y,\Theta^{(i-1)})\Sigma^{-1}\}]$$

Setting the derivative equal to null matrix and solving gives the complete-data ML estimate. Then the complete-data ML estimate is :

$$\hat{\Sigma} = \frac{1}{N}\sum\sum(\varepsilon_{ki}^{*}\varepsilon_{ki}^{*T}) + \frac{1}{N}\sum\sum(\Lambda_0\tilde{Z}_{ki}C_{\pi\varpi_{ki}}^{*}Z_{ki}^{*T}\Lambda_0{}^{T}) + \frac{1}{N}\sum\sum(\Lambda_0 Z_{ki}^{*}C_{\pi\varpi_{ki}}^{*T}\tilde{Z}_{ki}^{T}\Lambda_0^{T})$$
$$+\frac{1}{N}\sum\sum(\Lambda_0\tilde{Z}_{ki})D_{\pi}^{*}(\Lambda_0\tilde{Z}_{ki})^{T} + \frac{1}{N}\sum\sum(\Lambda_0 Z_{ki}^{*}D_{\varpi_{ki}}^{*}Z_{ki}^{*T}\Lambda_0^{T}) \tag{16}$$

where $\varepsilon_{ki}^{*} = y_{ki} - \Lambda_0\tilde{Z}_{ki}\pi^{*} - \Lambda_0\varpi_{ki}^{*}$

$$(5)\quad \frac{\partial Q(\Theta,\Theta_0)}{\partial \Lambda_0} = -[\sum\sum\Sigma^{-1}\Lambda_0\tilde{Z}_{ki}C_{\pi\varpi_{ki}}^{*}Z_{ki}^{*T}] - [\sum\sum\Sigma^{-1}\Lambda_0 Z_{ki}^{*}C_{\pi\varpi_{ki}}^{*T}\tilde{Z}_{ki}^{T}]$$
$$-[\sum\sum\Sigma^{-1}\Lambda_0\tilde{Z}_{ki}D_{\pi}^{*}\tilde{Z}_{ki}^{T}] - [\sum\sum\Sigma^{-1}\Lambda_0 Z_{ki}^{*}D_{\varpi_{ki}}^{*}Z_{ki}^{*T}]$$
$$+[\sum\sum\Sigma^{-1}y_{ki}\pi^{*T}\tilde{Z}_{ki}^{T}] + [\sum\sum\Sigma^{-1}y_{ki}\varpi_{ki}^{*T}Z_{ki}^{*T}] - [\sum\sum\Sigma^{-1}\Lambda_0\tilde{Z}_{ki}\pi^{*}\pi^{*T}\tilde{Z}_{ki}^{T}]$$
$$-[\sum\sum\Sigma^{-1}(\Lambda_0\tilde{Z}_{ki}^{*}\varpi_{ki}^{*}\pi^{*T}\tilde{Z}_{ki}^{T} + \Lambda_0\tilde{Z}_{ki}\pi^{*}\varpi_{ki}^{*T}Z_{ki}^{*T})] - [\sum\sum\Sigma^{-1}\Lambda_0 Z_{ki}^{*}\varpi_{ki}^{*}\varpi_{ki}^{*T}Z_{ki}^{*T}\Lambda_0^{T}]$$

then the ML estimates, $\hat{\Lambda}_0$, is:

$$\hat{\Lambda}_0 = [\sum\sum y_{ki}(\pi^{*T}\tilde{Z}_{ki}^{T} + \varpi_{ki}^{*T}Z_{ki}^{*T})]\times$$
$$[\sum\sum\tilde{Z}_{ki}(\pi^{*}\pi^{*T} + D_{\pi}^{*})\tilde{Z}_{ki}^{T} + Z_{ki}^{*}(\varpi_{ki}^{*}\varpi_{ki}^{*T} + D_{\varpi_{ki}}^{*})Z_{ki}^{*T}$$
$$+[Z_{ki}^{*}(\varpi_{ki}^{*}\pi^{*T} + C_{\pi\varpi_{ki}}^{*T})\tilde{Z}_{ki}^{T} + \tilde{Z}_{ki}(\pi^{*}\varpi_{ki}^{*T} + C_{\pi\varpi_{ki}}^{*})Z_{ki}^{*T}]^{-1} \tag{17}$$

Note that in deriving (17), we suppose that Λ_0 is a full matrix .

In sum the E-steps and M-steps are:

(1) E-step : Find $E\log[L(c,\Theta)\big|y,T_{\eta_j}^{(i-1)}]$, acting as if $T_{\eta_j}=T_{\eta_j}^{(i-1)}$,
M-step : Substitute the equation (13) with these quantities, and then we obtain new T_{η_j}, set $T_{\eta_j}^{(i)}$ equal to this new T_{η_j}.

(2) E-step : Find $E\log[L(c,\Theta)\big|y,T_{\eta_b}^{(i-1)}]$, acting as if $T_{\eta_b}=T_{\eta_b}^{(i-1)}$,
M-step : Substitute the equation (14) with these quantities, and then we obtain new T_{η_b}, set $T_{\eta_b}^{(i)}$ equal to this new T_{η_b},

(3) E-step : Find $E\log[L(c,\Theta)\big|y,T_{\pi}^{(i-1)}]$, acting as if $T_{\pi}=T_{\pi}^{(i-1)}$,
M-step : Substitute the equation (15) with these quantities, and then we obtain new T_{π}, set $T_{\pi}^{(i)}$ equal to this new T_{π}.

(4) E-step: Find $E\log[L(c,\Theta)\big|y,\Sigma^{(i-1)}]$,acting as if $\Sigma=\Sigma^{(i-1)}$,
M-step : Substitute the equation (16) with these quantities, and then we obtain new Σ, set $\Sigma^{(i)}$ equal to this new Σ.

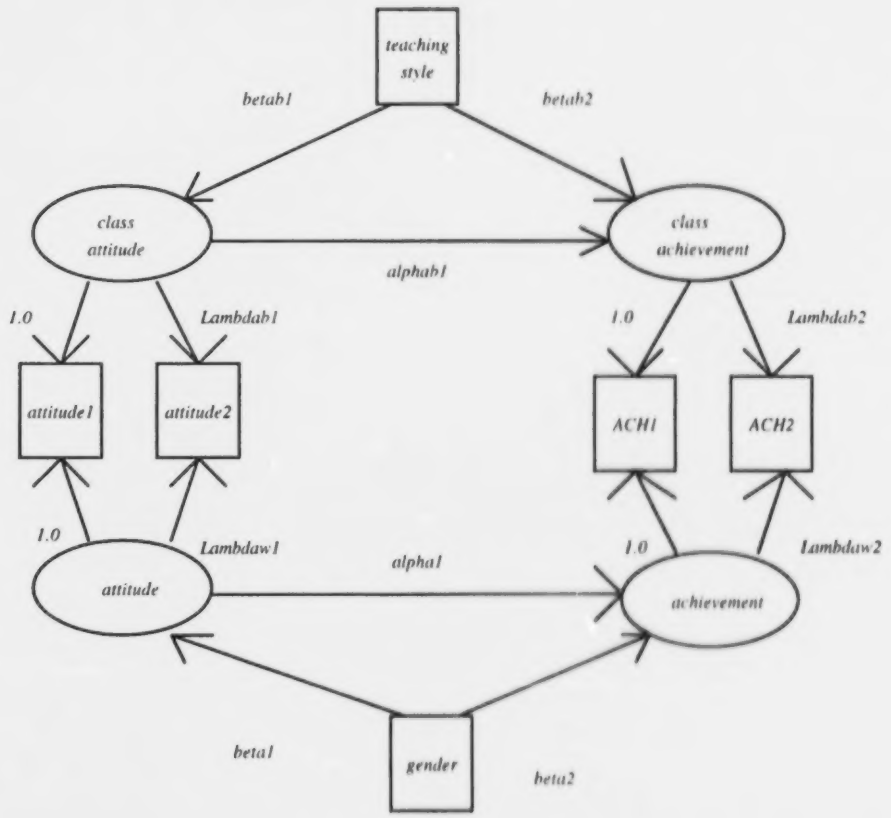
(5) E-step: After finding π^{*T}, ϖ_{kl}^{*T}, $D_{\varpi_{kl}}^{*}$, D_{π}^{*}, $C_{\pi\varpi_{kl}}^{*}$, acting as if $\Lambda_0=\Lambda_0^{(i-1)}$.
M-step: Substitute the equation (17) with these quantities, and then we obtain new Λ_0, set $\Lambda_0^{(i)}$ equal to this new Λ_0.

Then here the first iteration of the E and M step is completed. This algorithm proceeds until some user-specified termination criteria are met. For example, the algorithm might terminated when successive iterates differ from each other by no more than some number δ (i.e., $\delta=10^{-6}$). We conclude this section with the expression for the observed log-likelihood function. Convergence of the algorithm is monitored by computing the observed data log likelihood on each iteration, which is given by

$$\log L(\Psi,\Omega,\Lambda_0|y)=(N/2)\log(\det\Sigma)-(n_1/2)\log(\det T_{\eta_1})-,\ldots,-(n_J/2)\log(\det T_{\eta_J})$$
$$-(J/2)\log(\det T_{\eta_b})-(J/2)\log(\det T_{\pi})+\log(\det V_{11})+\sum\log(\det Q_{3j}^{-1})+\sum\log(\det U_j^{-1})$$
$$-\sum\sum[y_{ij}^{T}\Sigma^{-1}(y_{ij}-\Lambda_0\tilde{Z}_{ij}\pi^{*}-\Lambda_0 Z_{ij}^{*}\varpi_{ij}^{*})] \qquad (18)$$

4. An Example and Numerical Results

To illustrate how measurement and substantive theory can be integrated between and within levels in one overall framework, a hypothetical achievement model will be examined as an example. Consider a model where mathematics achievement is believed to be influenced by a student's gender and attitude toward mathematics and by a classroom characteristic teaching style. Teaching styles such as discovery-oriented instruction or expository teaching are believed to influence classroom mean attitude and achievement. Gender also is believed to be related to students' attitude and achievement on the individual level. The path-diagram for this hypothetical model is shown in Figure 1.



Fiigure 1. A Path Diagram for Multilevel Structural Equation Model

The table 2 is the result of fitting unbalanced data to the model. The total number of students is 1000. The focus of investigation is in discovering and testing the estimates of parameters are close to the predetermined population parameters within some what sampling error.

Table 1 Descriptive Statistics for Two Samples

	Mean	st. dev.	Sample covariance matrix			
Y1	0.122	5.203	50.122			
Y2	0.142	7.321	35.385	37.289		
Y3	0.175	5.734	13.291	12.982	54.937	
Y4	0.102	3.934	12.975	7.974	26.289	32.167

In discussing the results reported in Table 2, we say that the EM algorithm recovered the population parameters values fairly well. The criterion used for convergence of the observed log-likelihood is that the difference of (i-1) th log-likelihood and i-th log-likelihood is smaller than $(0.1)^6$ ($\delta \leq (0.1)^6$). In the 486 IBM PC with the speed of 66 mhz for the convergence it took about 1426th iterations.

Table 2. Parameter Estimates

$$\Lambda_0 = \begin{bmatrix} 1.0 & 0.0 & 1.0 & 0.0 \\ 0.75 & 0.0 & 0.75 & 0.0 \\ 0.0 & 1.0 & 0.0 & 1.0 \\ 0.0 & 0.70 & 0.0 & 0.60 \end{bmatrix}, \qquad \Lambda_0^{(0)} = \begin{bmatrix} 1.0 & 0.0 & 1.0 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 1.0 & 0.0 & 1.0 \\ 0.0 & 0.5 & 0.0 & 0.5 \end{bmatrix},$$

$$\hat{\Lambda}_0 = \begin{bmatrix} 1.0 & 0.0 & 1.0 & 0.0 \\ 076106 & 0.0 & 0.7443 & 0.0 \\ 0.0 & 1.0 & 0.0 & 1.0 \\ 0.0 & 0.72003 & 0.0 & 0.5928 \end{bmatrix}$$

$\Sigma = Diagonal(8.0, 10.0, 12.0, 14.0)$

$\Sigma^{(0)} = Diagonal(4.0, 5.0, 6.0, 7.0)$

$\hat{\Sigma} = Diagonal(7.935, 10.898, 11.785, 12.409)$

$$T_b = \begin{bmatrix} 20.0 & 10.0 & 0 & 0 \\ 10.0 & 30.0 & 0 & 0 \\ 0 & 0 & 20.0 & 8.0 \\ 0 & 0 & 8.0 & 20.0 \end{bmatrix}, \quad T_b^{(0)} = \begin{bmatrix} 10.0 & 5.0 & 0 & 0 \\ 5.0 & 15.0 & 0 & 0 \\ 0 & 0 & 10.0 & 4.0 \\ 0 & 0 & 4.0 & 10.0 \end{bmatrix},$$

$$\hat{T}_b = \begin{bmatrix} 17.942 & 11.28 & 0 & 0 \\ 11.28 & 27.94 & 0 & 0 \\ 0 & 0 & 19.44 & 8.042 \\ 0 & 0 & 8.042 & 21.601 \end{bmatrix}$$

We provide only the first 5 groups' T_j matrix.

$$T_1 = \begin{bmatrix} 25 & 10 \\ & 30 \end{bmatrix}, T_1^{(0)} = \begin{bmatrix} 10 & 5 \\ & 15 \end{bmatrix}, \hat{T}_1 = \begin{bmatrix} 27.16 & 8.083 \\ & 29.94 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 22 & 8 \\ & 28 \end{bmatrix}, T_2^{(0)} = \begin{bmatrix} 10 & 5 \\ & 15 \end{bmatrix}, \hat{T}_2 = \begin{bmatrix} 18.64 & 10.04 \\ & 27.15 \end{bmatrix}$$

$$T_3 = \begin{bmatrix} 18 & 7 \\ & 26 \end{bmatrix}, T_3^{(0)} = \begin{bmatrix} 10 & 5 \\ & 15 \end{bmatrix}, \hat{T}_3 = \begin{bmatrix} 18.23 & 7.18 \\ & 26.15 \end{bmatrix}$$

$$T_4 = \begin{bmatrix} 23 & 7 \\ & 25 \end{bmatrix}, T_4^{(0)} = \begin{bmatrix} 10 & 5 \\ & 15 \end{bmatrix}, \hat{T}_4 = \begin{bmatrix} 22.52 & 7.43 \\ & 26.83 \end{bmatrix}$$

$$T_5 = \begin{bmatrix} 21 & 8 \\ & 25 \end{bmatrix}, T_5^{(0)} = \begin{bmatrix} 10 & 5 \\ & 15 \end{bmatrix}, \hat{T}_5 = \begin{bmatrix} 20.57 & 8.52 \\ & 26.21 \end{bmatrix}$$

5. Conclusion

Research in the field of social sciences provides various challenges. For example, the random assignment of students to a set of conditions is not realistic in most cases. Even an experimental setting (Lumsdaine, 1963) the outcomes may show intracluster correlations due to the fact that (1) students do not receive their instruction individually but in groups, (2) interactions exist between treatments and students. This situation makes the application of the linear structural equation models to the real data inappropriate. As Cronbach (1976) pointed out, by not recognizing the nature of hierarchical data, many studies in the field of education have used inappropriate analyses. The difficulty of analyzing data arising from two levels is in assessing the nature of multivariable relationships at both levels simultaneously. This paper has shown how multilevel structural equation models can be formulated for hierarchical data and how they can be analyzed by using empirical Bayes with the EM algorithm.

One of the potential fields to which the mutilevel structural equation model may be extended is the model where the slope for the exogenous variables varies randomly across groups. Note that there are numerous settings in which multilevel structural equation models consisting of random slopes for exogenous variables are needed in order to represent adequately the variance-covariance structure of the data. Thus, for example, the gender gap in the SAT mathematics test scores can be explained by the group-level characteristics. The approach presented here can be generalized to incorporate random slopes.

As is well known, the EM algorithm is simple and numerically stable, but is slow. Recently Jamshidian and Jennrich (1993) developed a conjugate gradient scheme for accelerating the speed of the EM algorithm. In their AEM (accelerating EM) algorithm the evaluation of the gradient of the likelihood function is essential. When the number of parameters is moderate the implementation of the AEM

algorithm seems not so burdensome. To obtain the standard errors of the maximum likelihood estimates one may apply the SEM (supplemented EM) algorithm developed by Meng and Rubin (1991). However one can evaluate the observed Fisher information matrix using formulas provided by Raudenbush (1994). Future models can be expanded to larger and more complicated models. However many of the distributional and robustness properties of the model remain to be studied.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A, 149, 1-43.*

Bentler, P. M. (1989). *EQS : Structural Equation Program Manual*, Los Angeles, CA: BMDP Statistical Software, Inc.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin,* 101, 1, 147-158.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models in social and behavioral research: applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design and analysis.* Occasional paper of the Stanford Evaluation Consortium, School of Education, Stanford University.

Dempster, A. P., Laird, N. M & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B,* 39, 1-38.

Dempster, A. P., Rubin, D. B. & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association.*, 76, 341-353.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.

Goldstein, H. (1987). *Multilevel models in educational and social research.* C. Griffin & Co., London and Oxford University Press, New York.

Goldstein, H. (1989). Models for multilevel response variables with an application to growth curves. In R. D. Bock (Ed.), *Multilevel analysis of educational data (pp. 107-125*). New York: Academic Press, San Diego.

Graybill, F. A. (1983). *Matrices with applications in statistics.* Wadsworth Publishing Company, Inc. Belmont, California.

Jamshidian, M. & Jennrich, R. I. (1993). Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, 88, 221-228.

Joreskog, K. G. (1977). Structural equation models in the social sciences: specification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics.* Amsterdam: North-Holland.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL8 with PRELIS : A guide to the program and applications.* Chicago, IL: SPSS.

Lee, S. Y. & Poon, W. Y (1992). Two-level analysis of covariance structures for unbalanced designs with small level-one samples. *British Journal of Mathematical and Statistical Psychology.* 45, 109-123.

Lindley, D,V, & Smith, A.F.M. (1972) Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 1-41.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74, 817-827*

Longford, N. T. (1993). *Random Coefficient Models.* Clarendon Press, Oxford

Lumsdaine, A. A. (1963). Instruments and media of instruction. In N.L. Gage (Ed.),

Handbook of research on teaching. Rand Mcnally & Company, Chicago.
Mason, W. M., Wong, G. Y., & Entwistle, B. (1984). Contextual analysis through the multilevel linear model. *Sociological Methodology*, Sanfrancisco, CA: Jossey-Bass.
McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of mathematical and Statistical Psychology*, 42. 215-232.
Meng, X. L., & Rubin, D. B. 91991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical association*, 89, 899-909.
Muthen, B. O. (1988). *LISCOMP: User's guide*. Mooresville, IN:Scientific Software.
Muthen, B. O. (1990). *Means and covariance structure analysis of hierarchical data*. University of California at Los Angeles Statistics Series #62. Los Angeles: UCLA.
Muthen, B. O. (1994). Hierarchical Models of Ability at individual and class levels. *Intelligence 18, 165-187.*
Press, S. J (1982). *Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference*. Robert E. Krieger Publishing Company. Malabar, Florida.
Raudenbush, S.W. (1984). *Applications of a hierarchical linear model in educational research*. Unpublished doctoral dissertation, Havard University.
Raudenbush, S.W. & Bryk, A.S.(1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics*. 10, 75-98.
Raudenbush, S.W. (1988). Estimational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13, 85-116.
Raudenbush, S. W. (in press). Maximum likelihood estimation for unbalanced multilevel structural equation models via the EM algorithm. *British Journal of* Mathematical and Statistical Psychology.
Schmidt, W. H. (1969). *Covariance structure analysis of the multivariate random effects model*. Unpublished doctoral dissertation, University of Chicago.
Wu, C.F.J (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, 11, 95-103.